Automatic Seizure Detection in an In-Vivo Model of Epilepsy

Guillaume Saulnier and **Joelle Pineau**
School of Computer Science, McGill University,
Montreal, Quebec, Canada

Abstract

The goal of our research is to find patterns of EEG activity that will allow us to correctly identify seizures in living rats using machine learning techniques. Features are extracted from the EEG to characterize the signal over time. We perform model selection to reduce the set of features, as the goal is to have the algorithm running on a small personal device. The chosen features are used within a supervised classifier, based on randomized forests, in order to separate the different brain states. One of the challenges of this research is to detect all seizures, while preserving a low false positive rate, and low detection latency. We present results showing we can achieve this using data from three separate animals. The long-term goal of this research is to use this seizure detection method as part of a closed-loop adaptive neuro-stimulation device to reduce the incidence and duration of seizures.

Introduction

Epilepsy is a chronic neurological disorder characterized by spontaneous seizures. This disease affects roughly 50 million people worldwide, with 20% to 30% of the patients achieving no significant symptom reduction with standard pharmacological treatments (Loscher 1997). The causes are often unknown, varying from genetic pre-dispositions to head traumas, tumors, or infections.

The electroencephalogram (EEG), consisting of recording electrical signals from the brain, is often used to study the disease. EEG recordings provide large amounts of data which can be used to achieve automatic computational seizure detection.

The problem of automated seizure detection is the first step towards tackling many related problems, such as seizure prediction, prevention, and early seizure abortion. Automated seizure detection and prediction have been extensively studied in the context of human epilepsy (Mormann et al. 2007). Yet some of the more challenging problems, in particular seizure abortion and prevention using neuro-stimulation methods, are currently being studied in animal models of epilepsy (both *in-vitro* and *in-vivo*). Thus it is important to develop solid seizure detection techniques for these animal models also.

The work described below uses data collected from live (*in-vivo*) rats with chronic epileptiform activity. This animal model of epilepsy is highly isomorphic to human epilepsy, duplicating the disorder except for the aetiology (Curia et al. 2008).

The literature on automatic classification of seizures *in-vivo* is limited. Most detection procedures proposed to extract a single type of feature EGG; that feature is either used as input to a classifier, such as a neural network (Schuyler et al. 2007), or simply thresholded (Talathi et al. 2008). In general, the size of the moving windows on which features are computed is hand-selected. In cases where a classifier is used, it is trained on EEG recordings taken from different rats, and thus may not be a good fit for any of them.

As mentioned above, the literature on detection and prediction of human seizures is substantially more developed. A detailed survey indicates, amongst other findings, that seizure characteristics could vary substantially between patients (Mormann et al. 2007). This could be caused by the inherent heterogeny in either the type of epilepsy, the kind of manifestations or even treatments taken among individuals (Talathi et al. 2008). Thus, it was important to consider patient-specific classification criteria. The survey performed by Mormann et al. also observed that many previous work was limited, as in the *in-vivo* literature, to only a small set of features, often hand-selected, whereas many other features may be useful. Finally, the paper also raised the concern of poor statistical validation of performance in many earlier papers, and outlined a sound evaluation methodology.

Our approach to seizure detection tries to address these major concerns. First, we consider data collected on multiple animals, and optimize personalized classifiers using data from each animal. Multiple features (both single-channel and multi-channel) are computed from the EEG signal using a range of moving window sizes. An automated greedy feature selection is performed to select which type of feature and time resolution provide the best expected performance. This subset of features is passed to an ensemble classifier based on randomized forests. We select the classification thresholds using cross-validation within the training set, such as to simultaneously satisfy the key metrics: high detection accuracy, low false positive rate and low detection latency. Our method is validated using data from three separate animals, on which was recorded a total of 27 seizures.

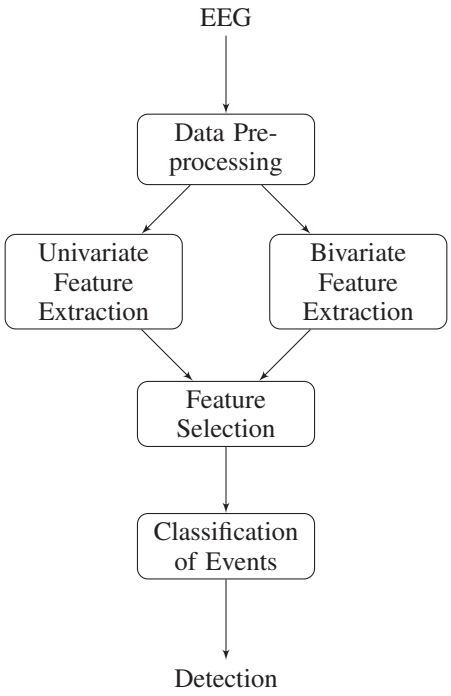

Figure 1: Flowchart describing the steps taken by the algorithm.

The methodology outlined in the paper may be useful for recognition and detection of dynamic events from a wide range of physiological sensors.

Methods

Figure 1 outlines the key steps taken by our algorithmic approach. Note that these steps are performed independently for each individual rat.

EEG Data and Preprocessing

A rat pilocarpine model of temporal lobe epilepsy was used to collect electroencephalographic data (Levesque et al. 2011). Status epilepticus was induced in three Sprague-Dawley rats (250-300 g.) by intraperitoneal injection of pilocarpine. Then, surgery was performed to place recording electrodes in the CA3 region of ventral hippocampus, the medial entorhinal cortex, the ventral subiculum, and the dentate gyrus. All procedures were approved by the Canadian Council of Animal Care and all efforts were made to minimize the number of animals used and their suffering.

Original recordings were sampled at 2000Hz, but downsampled to 200Hz without filtering. The dataset contained 27 seizures lasting an average of 79 ± 27 seconds. We extracted segments of approximately 5 minutes per seizure, such that the seizure is located in the middle, preceded and followed by non-seizure activity. These 27 seizures were observed between the 4^{th} and 15^{th} day after the pilocarpine injection. Only those with no artifacts nor signal saturation on all 4 electrodes were retained. The precise beginning and end of each seizure was annotated via spectral analysis of the EEG by an electrophysiologist. We also extracted 27 seizure-free segments of 5 minutes; these segments were chosen such that no seizures occurred in the preceding or following hour of recording. The selected segments were distributed between the three rats as outlined in Table 1.

Segments / Rat	Rat A	Rat B	Rat C
Seizure	6	7	14
Seizure free	6	7	14

Table 1: Distribution of Segments

Univariate Feature Extraction

Univariate features were computed on moving windows x of $\{1, 2, 5\}$ seconds, with $\{0, 1, 4\}$ seconds overlaps respectively, independently for each of the four channels of the EEG recordings. For each channel and window size, the following features were computed:

- The mean.
- The variance.
- The line length, defined to be the sum of the absolute difference of the amplitude of each consecutive pair of samples:

$$\sum_{i=1}^{N-1} |x(i) - x(i+1)|. \quad (1)$$

- Frequency components in the range $[1 - 100]$Hz, where a Hann window is first applied to the moving window to reduce edge effects.
- The mean of the convolution, performed in the frequency domain, with a finite impulse response filter equivalent to a Daubechies 4 wavelet at levels $\{0, 1, 2, 3, 4, 5\}$. See (Osorio, Frei, and Wilkinson 1998) for details.

Thus, 327 different feature points were created per channel, for a total of 1308 feature points per second of EEG recording.

Multivariate Feature Extraction

Multivariate features were computed on pairs of moving windows (x, y), such that x and y are extracted from different channels of the EEG recording. The moving window lengths considered were $\{1, 2, 5\}$ seconds, with $\{0, 1, 4\}$ seconds overlaps respectively, assuming always that x and y have the same length. The multivariate features were chosen based on those presented in (Mormann et al. 2007). Those can be split into two different categories, symmetric and non-symmetric:

Symmetric features (i.e. $f(x, y) = f(y, x)$):

- The Maximum Cross-Correlation, defined as:

$$C_{\max} = \max_{\tau} \left\{ \left| \frac{C_{xy}(\tau)}{\sqrt{C_{xx}(0) \cdot C_{yy}(0)}} \right| \right\}, \quad (2)$$

where

$$C_{xy}(\tau) = \begin{cases} \frac{1}{N-\tau}\sum_{i=1}^{N-\tau} x(i+\tau)y(i) & \tau \geq 0 \\ C_{xy}(-\tau) & \tau < 0 \end{cases}. \quad (3)$$

All $\tau \in [-0.5, 0.5]$ seconds were considered, as previously done in (Mirowski et al. 2009).

- The Maximum Cross-Correlation Index, defined as:

$$CI_{\max} = \arg\max_{\tau} \left\{ \left| \frac{C_{xy}(\tau)}{\sqrt{C_{xx}(0) \cdot C_{yy}(0)}} \right| \right\}, \quad (4)$$

where τ ranged as above.

Non-symmetric features (i.e. $f(x,y) \neq f(y,x)$):

- Linear Coherence, defined as:

$$\Gamma(f) = \left| \frac{G_{xy}(f)}{\sqrt{G_{xx}(f) \cdot G_{yy}(f)}} \right|, \quad (5)$$

where

$$G_{xy}(f) = FT_x(f) \cdot FT_x^*(f), \quad (6)$$

with $FT_x(f)$ denoting the Fourier transform of x at frequency f and $*$ the complex conjugate. The Linear Coherence was computed for $f \in \{10, 15, \ldots, 95\}$Hz.

- Two measures of non-linear interdependence, ${}^{x|y}S$ and ${}^{x|y}H$.
  Let $\{\vec{x}_i\}$ be the state space trajectory of $\{x_i\}$, where

$$\vec{x}_i = (x(i-(d-1)\tau), \ldots, x(i-\tau), x(i)), \quad (7)$$

with d being the dimension and τ the lag. We define α_{ij} and β_{ij} to be the time indices of the $j \in \{1, \ldots, k\}$ nearest neighbour of $\vec{x}_i$ and $\vec{y}_i$ in their state space, respectively. The two measures are:

$${}^{x|y}S = \frac{1}{N}\sum_{i=1}^{N} \frac{{}^{x}R_i^{(k)}}{{}^{x|y}R_i^{(k)}} \quad (8)$$

and

$${}^{x|y}H = \frac{1}{N}\sum_{i=1}^{N} \log \frac{{}^{x}R_i^{(N-1)}}{{}^{x|y}R_i^{(k)}}, \quad (9)$$

where

$${}^{x}R_i^{(k)} = \frac{1}{k}\sum_{j=1}^{k} (\vec{x}_i - \vec{x}_{\alpha_{ij}})^2 \quad (10)$$

and

$${}^{x|y}R_i^{(k)} = \frac{1}{k}\sum_{j=1}^{k} (\vec{x}_i - \vec{x}_{\beta_{ij}})^2. \quad (11)$$

The parameters used were $d = 10$, $\tau = 5$ ($\approx$ 23ms.) and $k = 5$, as proposed in (Mirowski et al. 2009).

- Three measures of phase synchronization, μ_{ps}, λ_{cp} and ρ_{se}.
  The phase of a channel x at time t, denoted $\phi_x(t)$, was extracted by convoluting the signal with a complex Gabor function $\mathcal{G}_{f,\alpha}(t)$:

$$\mathcal{G}_{f,\alpha}(t) = e^{-\alpha^2 t^2} e^{i2\pi f t} \quad (12)$$

yielding the convolved signal $W_x(t)$, from which the phase was extracted:

$$\phi_x(t) = \arctan \frac{Im(W_x(t))}{Re(W_x(t))}. \quad (13)$$

We converted $\mathcal{G}_{f,\alpha}(t)$ to a finite impulse response filter by assuming that $\forall t$ such that $e^{-\alpha^2 t^2} < 10^{-7}$, $\mathcal{G}_{f,\alpha}(t) = 0$.

– Mean phase coherence, defined as:

$$\mu_{pc} = \left| \frac{1}{N}\sum_{j=1}^{N} e^{i\phi_{x-y}(t_j)} \right|, \quad (14)$$

where

$$\phi_{x-y}(t_j) = \phi_x(t_j) - \phi_y(t_j) \quad (15)$$

is the phase difference between the two signals.

– Conditional probability index, defined as:

$$\lambda_{cp} = \frac{1}{L}\sum_{l=1}^{L} |r_l|, \quad (16)$$

where

$$r_l = \frac{1}{|M_l|} \sum_{\substack{j \\ \phi_x(t_j) \in M_l}} e^{i\phi_y(t_j)}, \quad (17)$$

with

$$M_l = \left\{ \phi_x(t_j) \middle| \phi_x(t_j) \in \left[\frac{l}{L}2\pi, \frac{l+1}{L}2\pi \right] \right\}. \quad (18)$$

L determines the number of bins present in $[0, 2\pi]$. Its value is set to be $L = e^{0.626+0.4\ln(N-1)}$ as defined in (Mormann et al. 2007).

– Shannon entropy index, defined as:

$$\rho_{se} = 1 + \frac{1}{\ln L}\sum_{l=1}^{L} p_l \ln p_l, \quad (19)$$

with

$$p_l = \frac{\left| \left\{ \phi_{x-y}(t_j) | \phi_{x-y}(t_j) \in \left[\frac{l}{L}2\pi, \frac{l+1}{L}2\pi \right] \right\} \right|}{|\{\phi_{x-y}(t_j)\}|}. \quad (20)$$

Multiple different parametrization of the Gabor function were used. We chose f and α such that 95% of the frequency response of $\mathcal{G}_{f,\alpha}$ was in the following ranges: $[0-4]$Hz, $[4-7]$Hz, $[7-13]$Hz, $[13-15]$Hz, $[14, 30]$Hz, $[30-45]$Hz and $[45-100]$Hz, again as defined in (Mirowski et al. 2009).

Thus, a total of 792 feature points per second of EEG recording were created.

Feature Selection

The computation of the full set of univariate and multivariate features $\mathcal{F}$ for an EEG is expensive. Therefore, we need to find a subset $S \subset \mathcal{F}$ that is small, but descriptive enough to obtain good detection performance.

The *receiver operating characteristic* (ROC) curve is commonly used for model selection. We calculate the area under the ROC curve, which we denote AUC, for each feature, using a training dataset. In the case of binary classification, the AUC is equivalent to the probability that the classifier will rank a randomly selected positive sample higher than a randomly selected negative sample (Fawcett 2006). The following three greedy selection schemes were used to selects subsets of $\mathcal{F}$.

I. S_{uni} contained the 30 univariate features with the highest AUCs.

II. S_{mixed} contained the 15 univariate and 15 multivariate features with the highest AUCs.

III. S_{each} was designed to select, for each *type* of feature, the parameterization (e.g. window length) with the highest AUC. The best Fourier transform parameters for each segment of 10Hz in the range $[1-100]$Hz were selected, yielding 10 features. As for the linear coherence, the computed frequencies ($\{10, 15, \ldots, 95\}$) were also cut into segments of 10Hz, from which we selected the best feature parameters, yielding another 9 features. The best parameters for the other 11 features (e.g. mean, variance, ..., Shannon entropy index) were selected, for a total of 30 features.

Feature vectors were then created by concatenating windows that end at the same time point for each feature present in the subset. Therefore, a feature vector containing 30 features exists for each second of EEG recording.

Classification of Events

We used a Forest of Extremely Randomized (Extra) trees (Geurts, Ernst, and Wehenkel 2006) to classify each feature vector. Each tree in the ensemble is built using the entire training set. At each node, K candidate tests are chosen randomly such that each contains an element of the feature vector and a random cut point. Then, a score based on the information gain is calculated for each candidate test. The best test is retained and the others are discarded. The data is split according to the test and children nodes are built. The process continues until either the split dataset becomes correctly separated (i.e. all feature vectors in the set are from the same class) or its size reaches a minimum $n_{\min}$. The label of a leaf is set to be the majority class of the split data set. We built a forest of $M = 200$ trees with $K = \lfloor\sqrt{|S|}\rfloor + 1 = 6$ and $n_{min} = \lfloor\log_2 |\text{Training set}|\rfloor + 1$.

To reach a final decision, each randomized tree in the forest is queried and the fraction of trees labeling the feature vector as seizure is reported by the forest. The forest output is then thresholded to separate the two classes: seizure or non-seizure. The threshold was chosen to be the middle point between the highest value for which all seizures were detected and the lowest value that did not cause any false alarms in seizure-free segments (considering again only the training set).

Evaluation

A 5-fold cross-validation was performed on each rat such that testing sets contained at least one seizure segment and one seizure-free segment. Feature selection and training of the classifier were performed using only the training data sets. The testing sets were used only to compute the following performance criterions:

i. Fraction of seizures detected: This is an indication of the quantity of seizures the algorithm successfully detected.

ii. Latency: The latency is the time between the start of a seizure and the first detection of the algorithm. If the seizure is missed, no latency is calculated. The average latency over the cross-validation is reported.

iii. Seizure-Free False Positives (SFFP): This number represents the quantity of seizure free segments for which at least one alarm was raised.

Given that we view the task of seizure detection within the larger therapeutic context, it is important to consider the setting within which the classifier will be used when selecting the performance criteria. In the context of a deep brain stimulation device designed to cause early seizure abortion based on detection events, a lower latency and higher seizure detection rate will be preferred, where as the cost of SFFP will be less important (presuming low side effect burden from the stimulation, which seems to be consistent with current devices). On the other hand, if the classifier is to be used to alert a third party that the patient is suffering from a seizure, such that help can be provided for the recovery, a low SFFP rate will be prioritized over a low latency. In general, we can trade-off between these different metrics by selecting the classification threshold appropriately. In our case, since the classifier was built for the general task of seizure detection, we optimized it for the best overall performance (i.e. the highest detection rate with the lowest SFFP and latency possible).

Performance on rat A

$S \subset \mathcal{F}$	Seizure detection	Latency (s.)	SFFP
S_{uni}	7 / 7	3.71 ± 5.44	0 / 7
S_{mixed}	7 / 7	3.57 ± 5.32	0 / 7
S_{each}	7 / 7	5.86 ± 5.49	0 / 7

Performance on rat B

$S \subset \mathcal{F}$	Seizure detection	Latency (s.)	SFFP
S_{uni}	6 / 6	3.00 ± 1.26	0 / 6
S_{mixed}	6 / 6	4.83 ± 2.23	0 / 6
S_{each}	6 / 6	5.50 ± 2.07	0 / 6

Performance on rat C

$S \subset \mathcal{F}$	Seizure detection	Latency (s.)	SFFP
S_{uni}	13 / 14	6.77 ± 6.92	1 / 14
S_{mixed}	13 / 14	7.62 ± 7.11	1 / 14
S_{each}	13 / 14	8.08 ± 7.30	0 / 14

Table 2: Performance criterions for each rat using personalized classifiers.

Results

Table 2 shows the performance criterions of the algorithm for each rat, with all three different feature selection

Performance on rat B

$S \subset \mathcal{F}$	Seizure detection	Latency (s.)	SFFP
S_{uni}	5 / 6	6.20 ± 5.26	0 / 6
S_{mixed}	5 / 6	25.4 ± 5.73	0 / 6
S_{each}	5 / 6	9.40 ± 2.07	0 / 6

Table 3: Performance criterions on rat B using a generalized classifier that was trained on data from all rats.

schemes. We observe that the latency using the subset of features S_{uni} is generally lower compared than the one using S_{mixed} or S_{each}. The latency tends to be greatest when selecting features of different types (S_{each}).

Rat C is the only one for which seizures were missed and SFFP occurred. The same seizure and the same seizure-free segment caused problems for both S_{uni} and S_{mixed}. Using S_{each}, a different seizure was missed, but the classifier produced no SFFP. It is important to note that even if the seizures were missed by the overall classifier, the forest still captured some amount of information as the fraction of trees labeling the seizure correctly reached up to 60%. A different choice of threshold would have resulted in the correct classification of all seizures, albeit at the cost of more SFFP.

It is well documented that seizures differ substantially between individuals. For this reason, our current method prioritizes personalized detection techniques. As we can see in Table 3, when we train a general classifier using EEG recordings from all the three rats, the performance diminishes. We find for rat B that a seizure is missed and the latency increases significantly. Yet it is possible that some aspects of the process, for example some of the feature selection, may generalize or transfer between individuals. A deep investigation of these questions is ongoing.

Conclusion

This paper outlines an algorithmic approach for automatic seizure detection. We discuss many of the methodological issues that arise when designing such a system, including feature selection, classifier training, and evaluation and interpretation of results. The approach is validated in the context of detecting seizures in living rats, where we show good performance on three main criteria: detection accuracy, latency, and false alarms.

Acknowledgments

The authors gratefully acknowledge the participation of Massimo Avoli and Maxime Levesque, of the Montreal Neurological Institute, for the collection and labeling of the EEG data used in the analysis. Financial support for this work is provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR).

References

Curia, G.; Longo, D.; Biagini, G.; Jones, R. S.; and Avoli, M. 2008. The pilocarpine model of temporal lobe epilepsy. *Journal of Neuroscience Methods* 172(2):143 – 157.

Fawcett, T. 2006. An introduction to roc analysis. *Pattern Recognition Letters* 27(8):861–874. ROC Analysis in Pattern Recognition.

Geurts, P.; Ernst, D.; and Wehenkel, L. 2006. Extremely randomized trees. *Machine Learning* 36(1):3–42.

Levesque, M.; Bortel, A.; Gotman, J.; and Avoli, M. 2011. High-frequency (80-500 hz) oscillations and epileptogenesis in temporal lobe epilepsy. *Neurobiology of Disease* In Press, Accepted Manuscript:–.

Loscher, W. 1997. Animal models of intractable epilepsy. *Progress in Neurobiology* 53(2):239 – 258.

Mirowski, P.; Madhavan, D.; LeCun, Y.; and Kuzniecky, R. 2009. Classification of patterns of eeg synchronization for seizure prediction. *Clinical Neurophysiology* 120(11):1927–1940.

Mormann, F.; Andrzejak, R. G.; Elger, C. E.; and Lehnertz, K. 2007. Seizure prediction: the long and winding road. *Brain : a journal of neurology* 130:314–333.

Osorio, I.; Frei, M. G.; and Wilkinson, S. B. 1998. Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. *Epilepsia* 39:615–627.

Schuyler, R.; White, A.; Staley, K.; and Cios, K. 2007. Epileptic seizure detection. *IEEE Eng Med Biol Mag* 26(2):74–82.

Talathi, S. S.; Hwang, D.; Spano, M. L.; Simonotto, J.; Furman, M. D.; Myers, S. M.; Winters, J. T.; Ditto, W. L.; and Carney, P. R. 2008. Non-parametric early seizure detection in an animal model of temporal lobe epilepsy. *Journal of Neural Engineering* 5:85–98.